From image descriptions to perceived sounds and sources in landscape: Analyzing aural experience through text

Chesnokova, Olga ; Purves, Ross S

Abstract: The importance of perception through all the senses has been recognized in previous studies on landscape preference, but data on aural perception, as opposed to the visual, remains rare. We seek to bridge this gap by analyzing texts that describe more than 3.5 million georeferenced images, created by more than 12000 volunteers in the Geograph project. Our analysis commences by extracting and automatically disambiguating descriptions that potentially contain verbs and nouns of sound (e.g. rustle, bellow, echo, noise) and adjectives of sound intensity (e.g. deafening, quiet, vociferous). Using random forests we classify more than 8000 descriptions based on the type of sound emitter into geophony (e.g. rustling wind, bubbling waterfall), biophony (e.g. gulls calling, bellowing stag), anthrophony (e.g. roaring jets, rumbling traffic) and perceived absence of sound (e.g. not a sound can be heard) with a precision of 0.81. Further, we additionally classify these descriptions as negative, neutral and positive using an Opinion Lexicon and GloVe word embeddings. Our results show that sentiment classification gives an additional level of understanding of descriptions classified into different types of sound emitters. We see that geophony, biophony and anthrophony cannot be uniquely classified as positive or negative. Our results demonstrate how text can provide a valuable, complementary to field-based studies, source of spatially-referenced information about aural landscape perception.

DOI: https://doi.org/10.1016/j.apgeog.2018.02.014

Originally published at:
Chesnokova, Olga; Purves, Ross S (2018). From image descriptions to perceived sounds and sources in landscape: Analyzing aural experience through text. Applied Geography, 93:103-111.
DOI: https://doi.org/10.1016/j.apgeog.2018.02.014

Applied Geography 93 (2018) 103–111

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

From image descriptions to perceived sounds and sources in landscape: Analyzing aural experience through text

Olga Chesnokova*, Ross S. Purves

Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

ARTICLE INFO

Keywords:
Sound experiences
User generated content
Natural language processing
Landscape

ABSTRACT

The importance of perception through all the senses has been recognized in previous studies on landscape preference, but data on aural perception, as opposed to the visual, remains rare. We seek to bridge this gap by analyzing texts that describe more than 3.5 million georeferenced images, created by more than 12000 volunteers in the Geograph project. Our analysis commences by extracting and automatically disambiguating descriptions that potentially contain verbs and nouns of sound (e.g. rustle, bellow, echo, noise) and adjectives of sound intensity (e.g. deafening, quiet, vociferous). Using random forests we classify more than 8000 descriptions based on the type of sound emitter into geophony (e.g. rustling wind, bubbling waterfall), biophony (e.g. gulls calling, bellowing stag), anthrophony (e.g. roaring jets, rumbling traffic) and perceived absence of sound (e.g. not a sound can be heard) with a precision of 0.81. Further, we additionally classify these descriptions as negative, neutral and positive using an Opinion Lexicon and GloVe word embeddings. Our results show that sentiment classification gives an additional level of understanding of descriptions classified into different types of sound emitters. We see that geophony, biophony and anthrophony cannot be uniquely classified as positive or negative. Our results demonstrate how text can provide a valuable, complementary to field-based studies, source of spatially-referenced information about aural landscape perception.

1. Introduction and background

What is the contribution of sounds to the way people perceive landscapes? And how can we gather information about such perceptions over large spatial scales? User Generated Content (UGC) has proven to be a suitable source for research questions dealing with such phenomena as people's perception of sense of place (Jenkins, Croitoru, Crooks, & Stefanidis, 2016), conceptualizations of natural features (Derungs & Purves, 2016), olfactory perception (Quercia & Schifanella, 2015), visual perception of landscapes (van Zanten et al. 2016) and assessment of the collective value of protected areas (Levin, Mark, & Brown, 2017). In this study we investigate another subjective phenomenon, namely aural perception of landscapes in UGC, with the underlying future aim of integrating sound information in landscape preference models.

Aural perception is an important constituent in landscape preference assessment (Brown & Brabyn, 2012; Sherrouse, Clement, & Semmens, 2011; Tudor, 2014) and is typically integrated using field surveys (Pilcher, Newman, & Manning, 2009) or laboratory sessions (Benfield, Bell, Troup, & Soderstrom, 2010; Manyoky, Wissen Hayek, Heutschi, Pieren, & Grêt-Regamey, 2014). However, these methods do not allow large regions to be characterized and are time consuming. We assume that aural perception of landscape is present in some written descriptions associated with photographs uploaded by individuals in UGC since photographs have been argued to be a good source of information related to shared experiences of places (Fisher & Unwin, 2005), and sound is one important element of such experiences. The following example vividly illustrates such use of language at an individual level: “If you press your nose to the computer screen, you might just catch the scent of the wild garlic, and if you listen carefully you should hear the song of willow warbler and blackcap.[1]” However, if we wish to analyze such descriptions, then important questions remain with respect to how they can be extracted, how common they are, and what properties they have.

1.1. Sound experiences

Although our sensory experience of nature is by definition multisensory, the visual is often privileged in both research and policy. Thus, despite the introduction of ‘soundscape’, ‘acoustic ecology’ and

* Corresponding author.
E-mail addresses: olga.chesnokova@geo.uzh.ch (O. Chesnokova), ross.purves@geo.uzh.ch (R.S. Purves).

[1] http://www.geograph.org.uk/photo/824881.

https://doi.org/10.1016/j.apgeog.2018.02.014
Received 22 December 2017; Received in revised form 20 February 2018; Accepted 21 February 2018
Available online 05 March 2018
0143-6228/

'soundscape ecology' (Southworth, 1969; Schafer, 1993; Pijanowski, Farina, Gage, Dumyahn, and Krause, 2011), aural perception is often of secondary importance in modelling landscape preferences. To relate sound to landscape preference it is important to consider the influence of perceived sound emitters as natural or unnatural (Fisher, 1999), rather than simply decibel values, since we do not hear abstract sounds, but "we hear the way *things sound*" (p. 40 Morton, 2009). Krause (2008), in collaboration with Gage, developed a useful taxonomy for sound emitters in landscape, identifying geophony (non-biological natural sounds), biophony (sounds produced by animals) and anthrophony (human-generated sounds).

Fisher (1999) claims that as soon as we perceive a sound as natural it has a positive aesthetic quality. Thus, similar sounds when perceived as being emitted by a jet engine or a waterfall would be considered unpleasant or "majestically powerful," respectively (p. 28–29 Fisher, 1999). Carles, Barrio, and De Lucio (1999) in their study of sound influence on landscape value note that similar to findings in visual perception, water sounds are typically positively connoted. Furthermore, discordant scenes, for example with positive visual (e.g. a water body) and negative aural cues (e.g. the sound of a busy road) were considered to be especially disturbing. In a series of soundwalks reported on by Pérez-Martínez, Torija, and Ruiz (2018), visitors characterized the sounds of certain emitters as being unpleasant, with, for instance, bird calls dominating, and thus detracting from landscape aesthetics. The negative effects of anthrophony are reported by Pilcher et al. (2009) to be especially important in wild areas, natural parks and other protected areas, where the intrusion of anthropogenic sounds is more disturbing. All of these studies provide us with useful clues as to how aural perception influences landscape perception, but none of them are easily applied across large regions.

1.2. User generated content and extraction of subjective phenomena from language

Our starting point is the hypothesis, based on an initial exploration of content, that UGC can be used to estimate aural perception of landscapes in the British Isles. This hypothesis is supported by previous work which has shown that, for example, tags associated with Flickr images or Tweets content have strong associations with place (Jenkins et al. 2016; Rattenbury, Good, & Naaman, 2007) or that olfactory perception of urban landscapes can be explored through UGC (Quercia & Schifanella, 2015). The same team of researchers also generated maps of urban noises using tags (Aiello, Schifanella, Quercia, & Aletta, 2016) by relating particular terms (e.g. church, car, dog) to particular sounds. However, their study implicitly links sounds to terms without clear evidence of the actual perception of sounds at a location. Similarly, analysis of spectrograms recorded by acoustic sensors (e.g. Pijanowski, Villanueva-Rivera, et al. 2011) does not allow a direct link between the presence of sounds and their perception by humans.

In this paper we build on previous work in two key ways. Firstly, the methods currently used in estimation of aural perception are time consuming and are not suitable for large regions. Using UGC provides an opportunity to explore the link between aural perception and landscapes across the British Isles. Secondly, in the case of recorded sounds presented in laboratory sessions the nature of a sound is abstracted from its context in the landscape. Therefore, we here set out to explore the efficacy of a range of methods for extracting and classifying textual descriptions related to aural perception of sounds, and apply sentiment analysis methods to explore the extent to which landscape descriptions related to different sound emitters can be characterized as positive, neutral or negative. We then explore, quantitatively and qualitatively how aural perception is characterized in our corpus, zooming in to explore local patterns in the description of sound experiences and zooming out to characterize the prominence and distribution of different sound experiences.

2. Data and methods

2.1. Data and study region

As a corpus we used descriptions associated with georeferenced pictures collated through the crowdsourced project Geograph British Isles. Geograph was launched in 2005 with the aim of documenting landscapes through the combination of representative pictures of a location and associated textual descriptions referring to individual grid squares at a granularity of 1 km in Great Britain and Ireland. Geograph contains simple game play elements, with the first contribution to a grid square being awarded more points, and has an active community of more than 12000 users. Similar to most UGC, contributions are biased, with a small number of users[2] contributing the majority of the data, but in previous work it has been shown that descriptions are not strongly biased by individual users, perhaps because of the clear aims and moderation of the uploaded photographs. Furthermore, in a survey carried out by the projects' initiators, users stated that it was important to be sure that the photographs and descriptions are archived for generations to come, and that they be used for educational purposes and promotion of local history. Since no mobile version of Geograph exists we assume that descriptions are written when photographs are uploaded from the desktop computer, though we found evidence that some users take notes in the field.[3] The data used in this paper were downloaded in June 2016, and consisted of more than 5 million photographs, of which more than 3.5 million also had a textual description, and are available under a Creative Commons Attribution-ShareAlike 2.5 License.

2.2. Method overview

Our approach to extracting, classifying and evaluating aural descriptions from the corpus involved three distinct methodological steps:

1. Extraction of descriptions referring to either experienced sounds or perceived absence of sound
2. Classification of the extracted descriptions according to a taxonomy of sound emitters
3. Allocation of sentiment values to each classified description of sound

Fundamental to our work in the first two tasks was the development of an annotated corpus, which was used to evaluate the quality of our extraction rules, and to serve as training and test data for our classifier. Fig. 1 gives an overview of the key steps carried out and described below.

2.2.1. Rules of annotation

As is typical in work on natural language, we created an annotated dataset to, firstly, better understand the properties and use of language in our corpus, secondly, to provide training data for our classifier, and thirdly to evaluate the efficacy of our methods. The annotated dataset contained examples of either descriptions referring to perceived sounds (and thus, not *per se* all detectable sounds) or their perceived absence and we classified these examples according to the type of referenced sound emitter (Table 1).

Descriptions of the following cases were all annotated as related to sound experience:

- aural perception at the moment the photograph was taken, for

[2] Detailed demographic data about users are not available, but based on a survey carried out by the project initiators it appears that users are in general more likely to be over 50 and male.

[3] I made a note on the map that whilst photographing this, the larks were almost deafening! Source: http://www.geograph.org.uk/photo/902702.

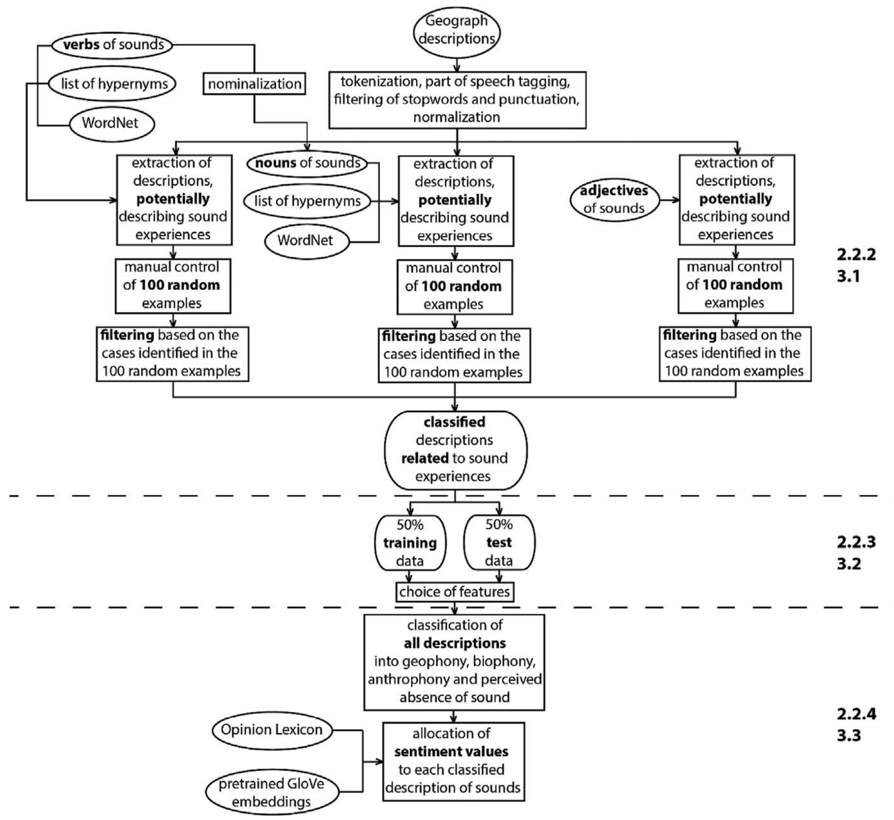

Fig. 1. Steps of data extraction and classification, relevant section numbers are indicated on the right.

Table 1
Types of sound emitters and their description after (Krause, 2008).

Type of sound emitter	Description
Geophony	Descriptions of natural sounds produced by non-biological sources, e.g. wind, waves, thunder, etc. ' … The pieces of ice were building up causing a swishing noise.[a]'
Biophony	Descriptions of sounds produced by animals. ' … If the picture came with sound, you'd hear the constant buzz of insects, the birds singing in the hedges and swifts screaming overhead. … [b]'
Anthrophony	Descriptions of sound produced by humans (including human voices) and anthropogenic objects (e.g. power plant). ' … Aircraft noise is a continual detractor in this intrinsically peaceful countryside.[c]'
Perceived absence of sound	Explicit description of absence of sound, e.g. 'quiet on Sunday morning,' 'not a sound can be heard.' ' … A curious, secret spot, yards away from the thunderous noise of the dual carriageway.[d]'
Mixed	Descriptions including two and more sound emitters, e.g. 'singing birds and roaring traffic' or ' … quiet canal, only the faint hum of the A1 can be heard. … [e]'
Unclear	The sound emitter is unclear, including the references to sound emitter as 'it' or 'they'. 'They look good, but they're noisy![f]'

[a] http://www.geograph.org.uk/photo/233383.
[b] http://www.geograph.org.uk/photo/3507826.
[c] http://www.geograph.org.uk/photo/2035700.
[d] http://www.geograph.org.uk/photo/572030.
[e] http://www.geograph.org.uk/photo/2012725.
[f] http://www.geograph.org.uk/photo/319060.

example, *skylarks are singing, running water can be heard*;

- motion of objects described using "sound verbs," e.g. *traffic thunders past*, *the stream gurgles*;
- explicit references to the possibility of sounds (even from the past): *apparently there is a marked echo in the area if one shouts loudly*;
- explicit references to the absence of sound: *the traffic no longer rumbles through their village*;
- aural perception expressed in poems included in the description; and
- indoor sounds.

Descriptions not classified as aural perception included the following:

- with no explicit reference to aural perception: *note the use of straw bales as a noise barrier*;
- of sounds produced by the author of the commentary (e.g. singing, whistling); and
- including similes or metaphors: *the blank walls cry out for some decoration.*

The full annotation was performed by only one person. To test the usefulness of the annotation rules, a second person annotated 100 randomly selected descriptions and inter-annotator agreement was calculated (Landis & Koch, 1977). For annotation of extracted (Cohens Kappa = 0.80) and classified sounds (Cohens Kappa = 0.88) inter-annotator agreement was *almost perfect* according to the classification of the Landis and Koch (1977), implying that annotation rules used are clear and that the annotation was consistent.

2.2.2. Extraction of descriptions related to sound experiences

We extracted descriptions of sound experiences using a combination of natural language processing methods. To reduce the effects of bias induced by participation inequality, we firstly removed similar descriptions generated by the same user by comparing sequences.

For all remaining descriptions we then carried out part of speech tagging, and using a lexicon of sound verbs extracted candidate sound descriptions after normalizing descriptions by lemmatization. Our initial list of verbs was based on those listed by Levin (1993) as verbs of sound emission, verbs of sounds made by animals and verbs of sound existence. To these verbs we added synonyms extracted from WordNet and clearly related to sound, leading to a total of 196 verbs. Since many of these verbs are polysemous we disambiguated verb usage at sentence level using WordNet hypernyms (categories) associated with the verb and its sentence context using the Lesk algorithm (Manning & Schutze, 1999). We carried out an analogous process for nouns after nominalizing our verb list. Finally, we also extracted descriptions using adjectives contained in a lexicon of sound-related adjectives. However, since WordNet does not contain adjectives in its hierarchy we manually reduced the lexicon of sound-related adjectives[4] to those we judged least likely to be used ambiguously (e.g. we retained *quiet* but not *pleasing*).

Since our rules aim at identifying candidate sound descriptions (i.e. high recall), we implemented them and then annotated a subset of candidate descriptions. Based on the properties of these subsets (i.e. commonly occurring false positives leading to lower precision) we then refined the rules used before annotating the sound descriptions extracted after refinement.

2.2.3. Feature choice

In order to classify descriptions related to sound experiences into types of sound emitter we use random forest classification.[5] Random forests are well suited to classification tasks using diverse feature types, are robust to extraneous features and are straightforward to train (Criminisi, Shotton, & Konukoglu, 2011). Very widely used features in training text classifiers are frequent n-grams – sequences of words found in text (i.e. unigrams are individual words, bigrams are sequences of two words) (Manning & Schutze, 1999). Since our classification task is to identify descriptions related to geophony, biophony, anthrophony and perceived absence of sound we use additional features we judged likely to be useful as described in Table 2. As well as these four classes, we also labelled (and thus trained our classifier on) descriptions belonging to mixed and unclear classes.

2.2.4. Sentiment analysis

The general procedure of allocating sentiment values to a text is described in (Iyyer, Manjunatha, Boyd-Graber, & Daumé , 2015) and was followed here. Firstly, we take an existing general Opinion Lexicon (Hu & Liu, 2004). Though it would be beneficial to have a domain-specific lexicon (Choi & Cardie, 2009), to our knowledge no such lexicon exists in the domain of landscape properties perception. Secondly, using a pretrained set of GloVe word embeddings (Pennington, Socher, & Manning, 2014) we train a gradient descent model[6] to assign polarity values (1 or -1) to all the words we have in our descriptions, and not only those contained in the Opinion Lexicon. Finally, we assign a sentiment value to each description by averaging word sentiment values for a description.

[4] http://www.sightwordsgame.com/parts-of-speech/adjectives/sound/.

[5] http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

[6] http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html.

3. Results and interpretation

3.1. Annotation and extraction of candidate sound descriptions

Annotation was carried out for all descriptions identified as candidate sound descriptions according to the rules described in §2.2.2. Table 3 gives a breakdown of this process, and we summarize important details below.

Based on our initial rulesets, we initially extracted 2436, 5247 and 11453 descriptions based on verbs, nominalized verbs and adjectives respectively. After filtering very similar descriptions and duplicates (i.e. descriptions extracted using our rulesets more than once) a total of 2250, 4730 and 11410 candidate descriptions remained.

For each set we then annotated 100 randomly selected descriptions, and calculated precision. We then used the false positives in each set of candidate descriptions to identify common errors and refined our rules on this basis. For verbs, our initial precision was 0.53. A small number of verbs appeared to be very commonly used polysemously (e.g. *echoes the style of Victorian buildings* or *the house was knocked down*). For a set of five such verbs, we then removed descriptions which contained only these, and no other sound verbs. After this refinement, we extracted 1653 descriptions and annotated all of these. Precision with our refined rules was 0.76.

For nouns, the initial precision was low (0.20) for an annotated random sample of 100 descriptions. A small number of very common polysemous nouns were removed if descriptions contained only these nouns (e.g. *the tree bark is very pretty* or *a clump of bushes is visible on the horizon*), and with the new rules we extracted 1342 descriptions with a precision of 0.68.

Adjectives generated by far the most descriptions, and based on an initial random sample of 100 descriptions precision was 0.56. Since we could not use the Lesk algorithm to disambiguate such adjectives (as hypernyms for adjectives are not contained in WordNet) ambiguity was not considered in our initial extraction. Exploring false positives we noted that *quiet* often appeared to be used in a more general sense to refer to frequency (e.g. *there is not much traffic on this quiet lane*), and added a rule to filter descriptions referring to quiet transport routes using a dependency parser.[7] We thus once again removed descriptions which only contained such phrases and triggered no other rules. Since, in contrast to our verbs and nouns, some 6805 descriptions were extracted, we annotated a random sample of 1000. Based on this sample we achieved a precision of 0.81 for descriptions extracted using adjectives.

After the process of annotation and extraction we created a final corpus of sound descriptions to be used as a training dataset in the classification step. For verbs and nouns we retained only those descriptions which we had annotated as containing sound, while for adjectives these were based on a precision of 0.81. Our complete collection thus contained 8784 descriptions contributed by 1074 unique users. 3036 of the descriptions were annotated.

Table 4 shows the classification of sound emitters according to our annotated corpus. Several points are worthy of note. Firstly, anthrophony is more common than either biophony or geophony. Secondly, mixed and unclear descriptions are relatively rare. Thirdly, geophony and anthrophony appear to be best extracted using a combination of verbs and nouns, while biophony is dominated by the use of verbs. In contrast, absence of sound is characterized by adjectives, reflecting that these descriptions emphasize a property of a location and are not in themselves captured by either verbs or nouns.

3.2. Data classification

Table 5 shows the results of a set of sensitivity tests exploring the

[7] https://spacy.io/usage/linguistic-features#section-dependency-parse.

Table 2
Features used in random forest classifier.

Feature	Description
1: Presence of frequent n-grams	N most frequent uni and bigrams from our corpus after removal of stop words and lemmatization (binary)
2: Presence of British birds and animals	- List of British birds (source: Wikipedia, 198 birds); list of British mammals (source: Wikipedia, 45 mammals) (binary)
3: Presence of transport related terms	Curated list of transport related terms (e.g. train, bus, railway, road, 14 terms) (binary)
4: Presence of natural landscape features and associated qualities	Selected elements based on the list of elements and qualities from Purves, Edwardes, and Wood (2011) (e.g. water, river, sea, hill, fog; 35 terms) (binary)
5: Frequency of references to classified roads	List of all classified roads identified using regular expressions of the form MXX, AXX and BXX where XX are 1 or more digits and M, A and B are motorways, primary and secondary routes (integer)

Table 3
Summary of the steps used to extract descriptions related to sound experiences.

Steps	Based on verbs	Based on nouns	Based on adjectives
Extraction using hypernyms and lists (only lists in the case of adjectives)	2436	5247	11453
After filtering very similar descriptions contributed by the same user	2250	4797	10817
After filtering descriptions already present in the previous dataset	–	4730	11410
Precision of 100 randomly selected examples	0.53	0.20	0.56
After filtering based on the results of the previous step	1653	1342	6805
New precision	0.76	0.68	0.81
Number of descriptions related to sound experiences	1265	909	862 annotated and 5748 unannotated

Table 4
Number of descriptions per type of sound emitter.

Type of sound emitter	Extracted using verbs	Extracted using nouns	Extracted using adjectives	Overall
Geophony	191	134	14	339
Biophony	355	110	60	525
Anthrophony	646	531	107	1284
Absence of sound	25	72	646	743
Mixed	29	41	19	89
Unclear	19	21	16	56
Total annotated	1265	909	862	3036

Table 5
Random forest classifier performance for different sound emitters and feature combinations.

Features (Table 2) Type	1 (200 unigrams)	1 (500 unigrams)	1 (500 unigrams), 2, 3, 4, 5	1 (500 unigrams), 2, 4
Geophony	P = 0.62	P = 0.84	P = 0.84	P = 0.86
	R = 0.34	R = 0.39	R = 0.31	R = 0.37
Biophony	P = 0.56	P = 0.66	P = 0.76	P = 0.74
	R = 0.54	R = 0.59	R = 0.67	R = 0.69
Anthrophony	P = 0.69	P = 0.72	P = 0.73	P = 0.74
	R = 0.81	R = 0.86	R = 0.90	R = 0.89
Absence of sound	P = 0.92	P = 0.91	P = 0.91	P = 0.92
	R = 0.86	R = 0.87	R = 0.86	R = 0.85
Overall	P = 0.70	P = 0.78	P = 0.81	P = 0.81
	R = 0.64	R = 0.68	R = 0.68	R = 0.70

contribution of various features to the classifier's overall performance, and also illustrates performance at the level of individual classes. Our classifier achieved best results (a precision of 0.81) using the 500 most common unigrams, our list of British birds and mammals and our list of natural features and related qualities. Adding transport related terms and named roads did not improve performance. Of note is the relatively high precision achieved for all classes (with values varying between 0.74 and 0.92) and the poor recall for geophony (0.37) implying that some two thirds of such instances were not identified. However, for this task we judge correct classifications (high precision) to be more important than high recall. Further, we concentrate on the four classes of geophony, biophony, anthrophony and absence of sound, because 98% of the descriptions belong to these classes.

Based on these results, we can map spatial distribution of classified sounds both for the whole corpus (Fig. 2) and explore descriptions of perceived sound experiences as extracted from Geograph locally (Figs. 3 and 4). With respect to Fig. 2 a few points are worthy of note. Firstly, the sound experiences extracted correlate with the overall distribution of images (Spearman rank, $r^2 = 0.67$). Secondly, they are dominated by absence of sound (5146) and descriptions of anthrophony (2275). Descriptions related to biophony (832) are less common, and least prevalent are those of geophony (386). These descriptions are also more prevalent in rural areas, and are only weakly correlated with the locations of anthropogenic sounds (Spearman rank, $r^2 = 0.10$ (biophony); $r^2 = 0.09$ (geophony)).

Fig. 3 demonstrates the efficacy of our approach for a rural area in Scotland, encompassing a range of scenic landscapes and a national park, but also traversed by important roads linking urban centers. Biophony is present in a number of descriptions of red deer, as well as the sounds of black grouse calling. Geophony is often related to water, especially *thundering* and *roaring* through gorges and over falls. Anthrophony is most often present in terms of traffic noise, especially where this is heard but not seen. Finally, despite the rural nature of the location, absence of sound, most often in terms of quiet is often reported. Fig. 4 shows results for an area of Central London. Here, geophony is absent completely, and biophony is reported only with respect to naturalized parrots in a park. There are a few references to anthrophony with respect to busy streets and a chiming clock, but the majority of detected references are to absence of sound. As in Fig. 3, these descriptions often contrast the scene with nearby surroundings, or make temporal comparisons (with the photograph taken at a quiet time).

3.3. Sentiment analysis

By calculating sentiment values for classified descriptions we can explore differences in the properties of descriptions, and potentially, the ways in which these are related to perceived environments. Almost 93% of words in our corpus (excluding stop words) were not contained in the Opinion Lexicon, demonstrating the importance of estimating sentiment values using pretrained word embeddings.

To illustrate the use of sentiment analysis in our sound descriptions we stratified sentiment values by generating three relative classes: a negative class consisting of all descriptions with a sentiment value more than half a standard deviation less than the mean, a neutral class of all

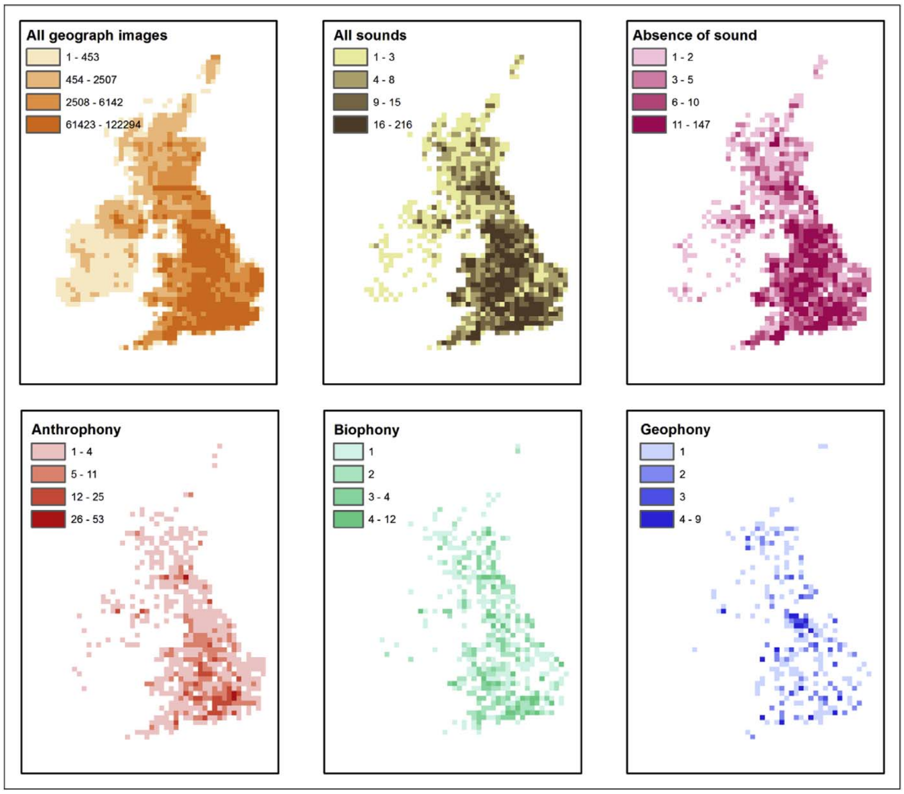

Fig. 2. Aggregated number of descriptions related to sound experiences per type of sound emitter.

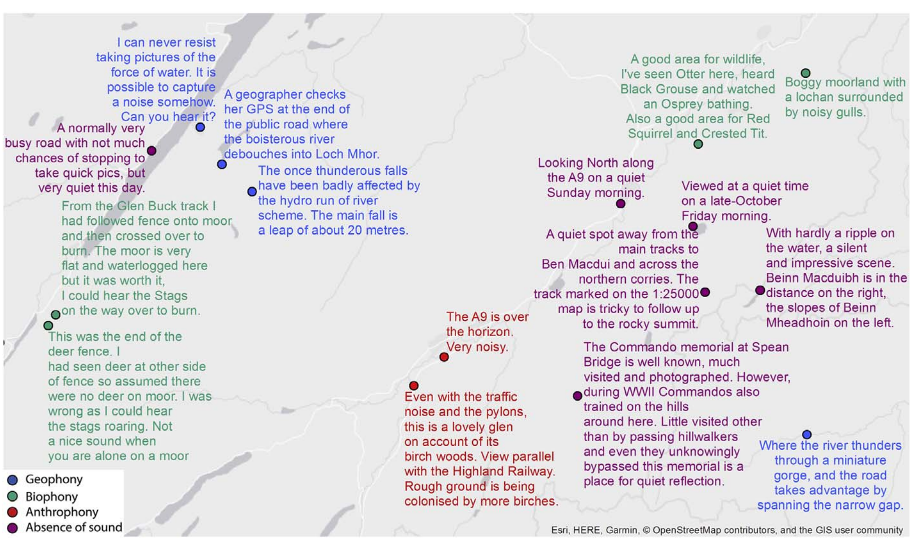

Fig. 3. An example of descriptions related to different types of sound emitter of the area between Loch Ness and Cairngorms National Park, Scotland, $n_{desc} = 16$, $n_{users} = 13$. Text source, associated images and authors: www.geograph.org.uk/photo/253824; www.geograph.org.uk/photo/1033986; www.geograph.org.uk/photo/1033836; www.geograph.org.uk/photo/1458692; www.geograph.org.uk/photo/593884; www.geograph.org.uk/photo/432524; www.geograph.org.uk/photo/680910; www.geograph.org.uk/photo/1773441; www.geograph.org.uk/photo/2940613; www.geograph.org.uk/photo/1055504; www.geograph.org.uk/photo/1582508; www.geograph.org.uk/photo/3206512; www.geograph.org.uk/photo/3088559; www.geograph.org.uk/photo/1580898; www.geograph.org.uk/photo/662610; www.geograph.org.uk/photo/1826916.

descriptions with sentiment values lying within half a standard deviation of the mean and a positive class consisting of the remaining descriptions with sentiment values greater than the mean plus half a standard deviation. Fig. 5 shows the distribution of descriptions as a function of their classification. Notable features include the strong association of geophony and biophony with negative descriptions (counter to our naïve expectations) and the association of absence of sound with neutral or positive descriptions. To explore the reasons for these distributions we generated word clouds of the 150 more frequently occurring terms for sentiment values.

Fig. 6 illustrates the resulting word clouds for geophony and biophony respectively. In the negative word clouds for geophony, many weather related words such as *thunder*, *rain*, *wind*, *gale* and *storm* are present. These were not present in the Opinion Lexicon, but have been assigned negative values due to their relationship with other words in the training data, presumably relating negative experiences to weather. *Thunder*, *rain* and *storm* are also prominent in the positive word cloud, along with other terms such as *rainbow*, *waterfall* and *sun*. Associated with negative biophony are many different animals and birds, together with *noise* and some types of sound emission (e.g. *hiss* and *bark*). Positive biophony is related to singing birds and wildlife, and appears, as for geophony, to be related to more natural terms associated with pleasant conditions.

However, we are also interested in how perception of sound varies within particular regions, and in Fig. 7 we explored absence of sound within the boundaries of the UK's 15 national parks. In these word clouds we only retained words which were unique to negative or positive sentiment. Negative sentiment with respect to absence of sound

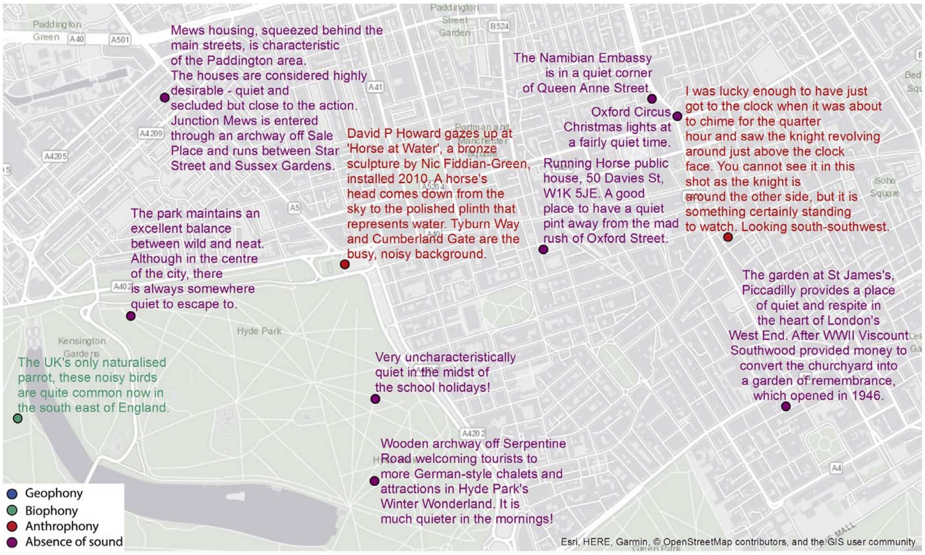

Fig. 4. An example of descriptions related to different types of sound emitter of London, England, $n_{desc} = 11$, $n_{users} = 10$.
Text source, associated images and authors: www.geograph.org.uk/photo/4760309, www.geograph.org.uk/photo/4646159; www.geograph.org.uk/photo/4270494; www.geograph.org.uk/photo/527664; www.geograph.org.uk/photo/2548274; www.geograph.org.uk/photo/1325858; www.geograph.org.uk/photo/1628770; www.geograph.org.uk/photo/119667; www.geograph.org.uk/photo/1999350; www.geograph.org.uk/photo/1661004; www.geograph.org.uk/photo/4418617.

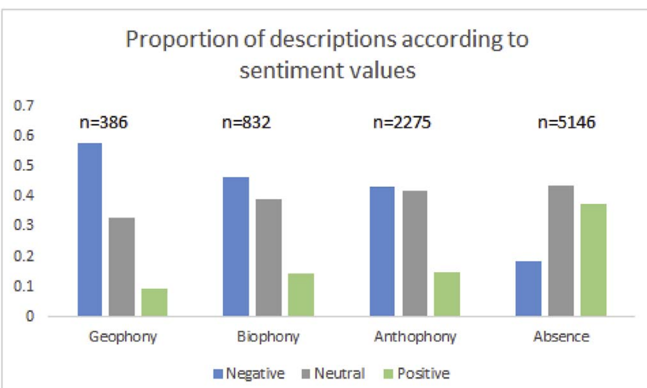

Fig. 5. Proportion of descriptions according to sentiment values.

appears to often be related to human activities (e.g. *pump*, *pub*, *railway*, *work*) as well as traffic and isolation (e.g. *traffic*, *backwater*). By contrast, terms relating to positive absence of sound often relate to positively connoted adjectives (e.g. *tranquil*, *enjoy*, *peaceful*, *attractive*, *lovely*) and contain more natural landforms (e.g. *beach*, *summit*, *bay*, *pass*).

4. Discussion

In the following, we discuss our results from two, contrasting, perspectives. Firstly, we explore our methodological contribution, setting out strengths and weaknesses of our approach to the extraction and classification of sound experiences, and the use of sentiment analysis methods on these descriptions. Secondly, we explore our results in the context of previous research on sound experiences both through traditional approaches in landscape research and research based on extraction of perception through UGC.

Our first important contribution is the creation of an annotated, classified corpus of sound descriptions consisting of 8784 descriptions associated with georeferenced images. Creating this corpus would not have been possible without the use of our heuristic methods to extract these descriptions, which were iteratively developed and have a mean precision of 0.75. Our heuristics, based on sound-related lexicons and, in the case of verbs and nouns, disambiguation using hypernyms and the Lesk algorithm are thus sufficiently accurate to allow us to reliably extract sound descriptions from a large collection. However, as is often the case in such work, we have no knowledge of the recall of our method, since this would require us to annotate by hand a very large volume of descriptions. In this particular case, because sound descriptions are rare (making up around 0.25% of our corpus in total) we would have to annotate some 400 descriptions to find a single sound description, and such a manual approach would be prohibitively time consuming. Further to creating our corpus of sound descriptions, we trained a classifier to allocate these to the classes from the taxonomy proposed by Krause (2008). Our best performing set-up used the 500 most frequent unigrams, presence of British birds or mammals and presence of natural features and related qualities in descriptions and achieved a precision of 0.81. However, here we were also able to estimate recall, since our classifier ran on annotated examples of sound experiences. Although overall recall was excellent (0.70) we note that in the case of geophony our classifier performed less well, with a recall of only 0.37. The most likely explanation for this poor performance is the low number of examples of geophony overall, resulting in limited training data for the classifier, especially when compared to anthrophony and absence of sound. However, it is important to note that our approach gives high precision – in other words though not all examples of geophony are classified, those that are, are typically correctly classified. To carry out sentiment analysis we used an Opinion Lexicon to assign values to every non-stop word in a description. Since only around 7% of the words in our descriptions were contained in the lexicon, we used word embeddings and a gradient descent model to assign polarities to the remaining 93% of words. It is important to note that the polarities in the original lexicon are based on general connotations of words with positive or negative polarity, and not those specific to landscapes. Thus, *wild, mystery* and *frozen* all have negative polarities, although all of these terms might be associated with positive values in landscape terms. For example, mystery is suggested as a predictor of environmental preference (Kaplan & Kaplan, 1989). Our approach demonstrates how sentiment analysis can be used to stratify aural descriptions, and as shown in Figs. 6 and 7, to generate interpretable summaries of some landscape properties in terms of sounds and preferences.

Methodologically, our approach has a number of limitations. Firstly, our methods have been developed on a specific collection, and although the rules are general, they have not been tested on other corpora. Nonetheless, by privileging precision over recall, we are reasonably confident that the approach taken should work on other, similar corpora. Secondly, our methods are dependent on annotated data, and annotating is challenging even for humans. Thus, despite good inter-annotator agreement, some cases, especially those describing silence and/or quietness are ambiguous with respect to whether the silence is

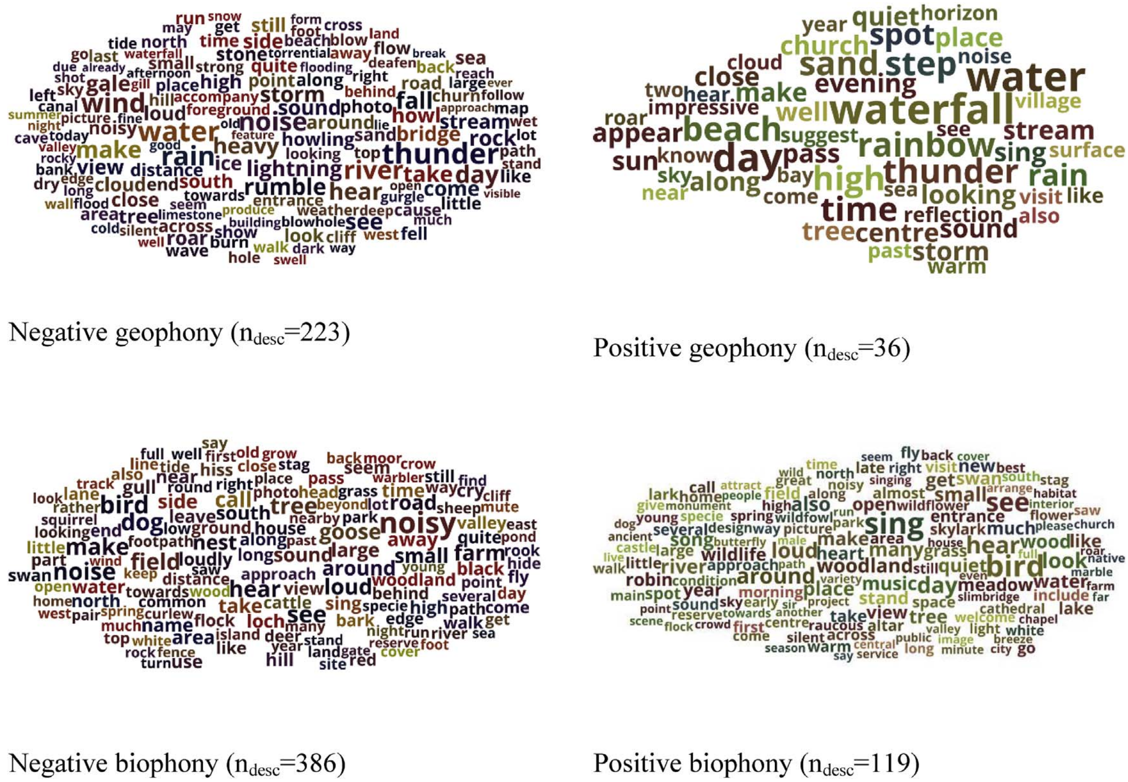

Fig. 6. 150 most frequent tokens occurring more than three times for negative and positive descriptions of geophony and biophony.

metaphorical or really related to sound. Thirdly, our approach to sentiment analysis is fairly simplistic, and in particular does not currently deal with negation or normalize according to the lengths of descriptions. And finally, we do not take into account dynamic nature of sounds varying over the course of a day or due to changing weather events.

Some of the strengths of our approach are well illustrated by exploring the content in more detail, and comparing it to previous research in this area. Since we generate a large corpus of classified descriptions this is possible, and as illustrated in Figs. 2–4 patterns can be explored both globally and locally. Our results appear to confirm that Krause's (2008) taxonomy is broadly useful, since we find abundant examples of anthrophony, biophony and geophony. However, it is important to note that in our work we take a human-centered approach as in Truax (1978) and Schafer (1993), where we do not concentrate on the whole variety of sounds in the environment but only on sounds perceived and recorded in written descriptions. Furthermore, exploring these patterns at local levels (e.g. Figs. 3 and 4) gives insights into how people perceive landscapes as a combination of all senses. Our methods thus provide a further potential approach in including sound in methods for landscape characterization (c.f. Sherrouse et al. 2011, Brown & Brabyn, 2012).

Sentiment analysis allowed us to stratify descriptions in meaningful ways. Thus, for geophony we observed that adverse weather seemed often to be related to negative descriptions, while positive descriptions, though relatively rare, seemed to be associated with landforms such as waterfalls and beaches, as well as positively connoted weather phenomena such as rainbows. In accord with previous work (e.g. Morton, 2009), our results appear to demonstrate the importance of perceived sound emitters, and we note that for biophony sounds associated with farm (and thus perceived less natural) animals appear to be less well received than wild animals. Further, we also observe that there are differences with respect to sounds related to wild animals, with for example gulls being perceived negatively while skylarks are positive. Similarly, we note that sounds which could be classified as *technophony* (Mullet, Gage, Morton, & Huettmann, 2016) in the anthrophony class appear to be mostly associated with negative sentiment (e.g. words such as *traffic*, *road*, *motorway*, *turbine* are common), while more positively

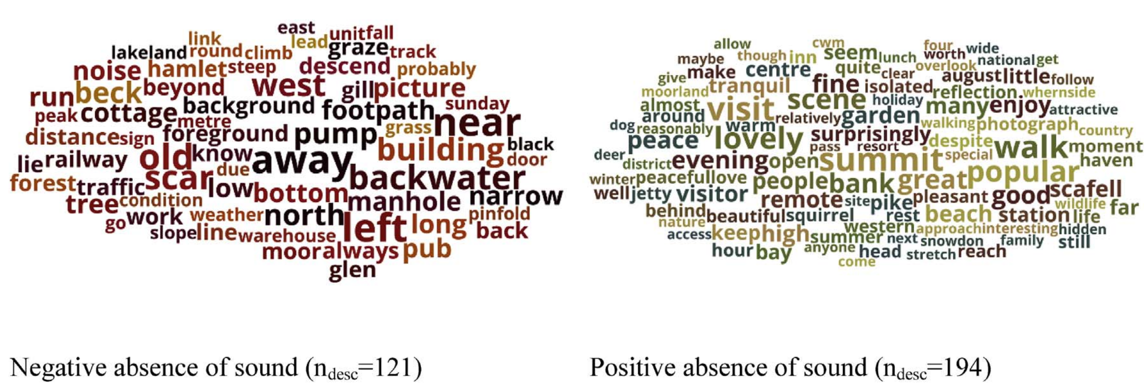

Fig. 7. Most frequent unique tokens occurring more than three times for negative and positive descriptions of absence of sound in the UK's 15 national parks.

associated anthrophony is reflected by words such as *clock*, *bell*, *music* and *sing*, combining *mechanistic* (e.g. chiming clock) and *oral* (e.g. carols singing) sub-classes of anthrophony (Qi, Gage, Joo, Napoletano, & Biswas, 2008). Again, these results are strikingly congruent with previous work (Pérez-Martínez et al. 2018), suggesting that our approach can usefully complement existing approaches to characterizing aural experiences.

5. Conclusions and outlook

Our aim was to explore the potential of textual descriptions associated with georeferenced photographs as a source of information on perceived sounds in landscapes. Since the dataset used was created by more than 12000 contributors, the resulting extracted descriptions provide us with a bottom-up view of the ways in which sounds are described, and give insights into how landscapes are perceived through multiple senses. Although the overall number is a small proportion of the corpus (some 0.25%), in absolute terms we have extracted more than 8000 sound-related descriptions, classified these according to emitters, and explored how the use of descriptions (and thus the perception of landscape in terms of sound) varies at different scales. Furthermore, by applying sentiment analysis we stratified descriptions and explored preferences within different classes of emitter, moving away from, for example, naïve expectations that natural sounds are *per se* positively evaluated.

Methodologically our contribution can be seen in two ways. Firstly, we have created an annotated corpus of classified descriptions which can serve as a basis for further research. Secondly, we have demonstrated how a combination of methods from natural language processing, going beyond simple extraction based on keywords, and taking account of typical linguistic phenomenal such as syntax and polysemy, allow us to extract and classify sound descriptions with high precision. Our approach to sentiment analysis used word embeddings to learn sentiment values for words not contained in our lexicon. Here we note that results are dependent on the lexicon used, and we propose to develop a domain-specific opinion lexicon focussed on landscape.

Our methods have general potential for future work in a number of ways. For example, they can be used to explore change in perceived sounds over time and thus contribute to the digital humanities. Furthermore, by exploring the relationship between aural descriptions and spatially contiguous models of abstract landscape qualities such as wilderness or tranquility the influence of perceived sounds on such properties can be accorded greater importance than is currently the case. Finally, we see great potential for integrating our results into a more general model of landscape preference based on textual analysis.

Acknowledgements

We are very grateful to Barry Hunter and Robin Stott for sharing anonymous results of the Geograph British Isles users' survey, and we would like to gratefully acknowledge all the contributors to Geograph British Isles (Creative Commons Attribution-ShareAlike 2.5 License). OC is also grateful to Lone Kristensen and Veerle Van Eetvelde for feedback on early versions of this work. Finally, we thank the reviewers for their useful suggestions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apgeog.2018.02.014.

References

Aiello, L. M., Schifanella, R., Quercia, D., & Aletta, F. (2016). Chatty maps: Constructing sound maps of urban areas from social media data. *Royal Society Open Science, 3*, 150690.

Benfield, J. A., Bell, P. A., Troup, L. J., & Soderstrom, N. C. (2010). Aesthetic and affective effects of vocal and traffic noise on natural landscape assessment. *Journal of Environmental Psychology, 30*(1), 103–111.

Brown, G., & Brabyn, L. (2012). The extrapolation of social landscape values to a national level in New Zealand using landscape character classification. *Applied Geography, 35*(1–2), 84–94.

Carles, J. L., Barrio, I. L., & De Lucio, J. V. (1999). Sound influence on landscape values. *Landscape and Urban Planning, 43*(4), 191–200.

Choi, Y., & Cardie, C. (2009). Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. *EMNLP '09 proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 590–598). .

Criminisi, A., Shotton, J., & Konukoglu, E. (2011). Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision, 7*(2–3), 81–227.

Derungs, C., & Purves, R. S. (2016). Characterising landscape variation through spatial folksonomies. *Applied Geography, 75*, 60–70.

Fisher, J. A. (1999). The value of natural sounds. *Journal of Aesthetic Education, 33*(3), 26–42.

Fisher, P., & Unwin, D. J. (2005). In D. J. Fisher, & Unwin (Eds.). *Re-presenting geographical information systems* (pp. 1–14). London: Wiley Sons London.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining KDD 04: Vol. 4*, (pp. 168–). .

Iyyer, M., Manjunatha, V., Boyd-Graber, J., & Daumé, H., III (2015). Deep unordered composition rivals syntactic methods for text classification. *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing*.

Jenkins, A., Croitoru, A., Crooks, A. T., & Stefanidis, A. (2016). Crowdsourcing a collective sense of place. *PLoS One, 11*(4), e0152932.

Kaplan, R., & Kaplan, S. (1989). *The experience of nature: A psychological perspective.* Cambridge University Press.

Krause, B. (2008). Anatomy of the soundscape. *Journal of the Audio Engineering Society, 56*(1/2).

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Levin, B. (1993). *English verb classes and alternations.* University of Chicago Press.

Levin, N., Mark, A., & Brown, G. (2017). An evaluation of crowdsourced information for assessing the importance of protected areas. *Applied Geography, 79*, 115–126.

Manning, C. D., & Schutze, H. (1999). *Foundations of statistical natural language processing.* The MIT Press.

Manyoky, M., Wissen Hayek, U., Heutschi, K., Pieren, R., & Grêt-Regamey, A. (2014). Developing a GIS-based visual-acoustic 3D simulation for wind farm assessment. *ISPRS International Journal of Geo-Information, 3*(1), 29–48.

Morton, T. (2009). *Ecology without nature: Rethinking environmental aesthetics.* Harvard: Harvard University Press.

Mullet, T. C., Gage, S. H., Morton, J. M., & Huettmann, F. (2016). Temporal and spatial variation of a winter soundscape in south-central Alaska. *Landscape Ecology, 31*(5), 1117–1137.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. *Conference on empirical methods in natural language processing (EMNLP 2014)*.

Pérez-Martínez, G., Torija, A. J., & Ruiz, D. P. (2018). Soundscape assessment of a monumental place: A methodology based on the perception of dominant sounds. *Landscape and Urban Planning, 169*, 12–21.

Pijanowski, B. C., Farina, A., Gage, S. H., Dumyahn, S. L., & Krause, B. L. (2011a). What is soundscape ecology? An introduction and overview of an emerging new science. *Landscape Ecology, 26*(9), 1213–1232.

Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., et al. (2011b). Soundscape Ecology: The science of sound in the landscape. *BioScience, 61*(3), 203–216.

Pilcher, E. J., Newman, P., & Manning, R. E. (2009). Understanding and managing experiential aspects of soundscapes at muir woods national monument. *Environmental Management, 43*(3), 425–435.

Purves, R. S., Edwardes, A. J., & Wood, J. (2011). Describing place through user generated content. *First Monday, 16*(9) Peer-Reviewed Journal on the Internet.

Qi, J., Gage, S. H., Joo, W., Napoletano, B., & Biswas, S. (2008). Soundscape characteristics of an environment: A new ecologival indicator of ecosystems health. In W. Ji (Ed.). *Wetland and water resource modeling and assessment* (pp. 201–211). New York: CRC Press.

Quercia, D., & Schifanella, R. (2015). Smelly Maps: The digital life of urban smellscapes. *9th international AAAI conference on web and social media*.

Rattenbury, T., Good, N., & Naaman, M. (2007). Towards automatic extraction of event and place semantics from Flickr tags. *Proc.of SIGIR: 07*, (pp. 103–110).

Schafer, R. M. (1993). *The soundscape: Our sonic environment and the tuning of the world.* Inner Traditions/Bear & Co.

Sherrouse, B. C., Clement, J. M., & Semmens, D. J. (2011). A GIS application for assessing, mapping, and quantifying the social values of ecosystem services. *Applied Geography, 31*(2), 748–760.

Southworth, M. (1969). *The sonic environment of cities. Environment and behavior* (June) 49–70.

Truax, B. (1978). *Handbook for acoustic ecology.* Burnaby, B.C. Canada: ARC Publications.

Tudor, C. (2014). *An approach to landscape character assessment* (October)56.

van Zanten, B. T., van Berkel, D. B., Meetemeyer, R. K., Smith, J. W., Tieskens, K. F., & Verburg, P. H. (2016). Continental scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences,* 1–7.